Benford's Law in Digital Pathology

Jasmine Caballero MD[1], M. Cynthia Duran MD[1], Mali Salmon-Divon PhD[2]
jasmine.a.caballero@gmail.com, cyn.deeduran@gmail.com, Universidad Autónoma de Guadalajara[1], Ariel University[2]

Objectives

- Show that histological normal cells hold a naturalness that correlates with Benford's Law
- Expand the analysis tool used on histology slides in digital pathology to include Benford's Law
- Open the possibility of being able to differentiate normal cells from mutated cells down to a 1% variance

Introduction

Cancers are classified by unregulated growth and the ability to spread. 70% of all cancers start in epithelial cells. Normal cells experience three different types of alterations when exposed to stress: atrophy, hypertrophy, and hyperplasia. Cancer can be seen by the naked eye when it reaches one billion cells. This is where digital pathology steps in. Digital pathology allows cells to be analyzed at a resolution higher than the human eye can see. Currently, the method used for detecting the differentiation of normal cells from cancer cells is known as the deep learning approach.

The field of image analysis has also been expanding along with digital pathology. One of the laws of statistics that is used to measure naturalness is Benford's Law. Benford's Law has been used in various fields such as: tax fraud, voting fraud, sports, music, sizes of volcanoes, bacterial populations, and radioactive decay. Benford's Law has yet to be utilized for image analysis in histology. This could lead to individualized, tailor-based medicine.

Benford's Law

Benford's Law was first observed by Simon Newcomb in 1881 and reinforced by Frank Benford in 1938. When a unit of distribution is categorized numerically, the natural count of the frequency of each unit {1,2,.....,9} is also known as first digit distribution. This has been found true for many forms of natural variations.

Benford's Law Cont.

A set of numbers satisfies Benford's Law for the Leading Digit if the probability of observing a first digit of d is log10(d+1/ d). In order for this to hold, solutions must be infinite. Since it is a finite unit that is being observed we must approximate, as a good visual fit. Histology cells are not expanded with first digits from {1-9} therefore it can not accurately follow Benford's Law. With digits that do not satisfy the first law, a Second Digit probability was created. A data set satisfies Benford's Law for the Second Digit if the probability of observing the second digit is approximately the generalized distribution of Benford's second order or Log10 (n+1)-Log10 (n)= Log10(1+1/n).

Digit	1st digit	2nd digit
0		0.119679
1	0.301030	0.113890
2	0.176091	0.108821
3	0.124939	0.104330
4	0.096910	0.100308
5	0.079181	0.096677
6	0.066947	0.093375
7	0.057992	0.090352
8	0.051153	0.087570
9	0.045757	0.084997

Materials and Methods

Histology slides were obtained from Penn State Bio-Atlas, Human Atlas. Image J was used to collect 500 diameters and areas then listed in excel, as the observed values. The distribution of the second digits were calculated using excel. The number of the expected digits was found using Benford's probability of second digits multiplied by the number of cells sampled. The observed frequency was analyzed in comparison to the expected probability of each digit {1,2….9}. To test if the second digit distribution of the areas or diameter conforms to Benford's Law, we used Pearson Chi Square Goodness-of-Fit test.

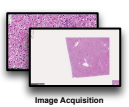


Image Acquisition

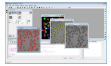
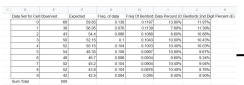


Image J Area/ Diameter Analysis


Second Order Digit Extracted and Compared to Expected

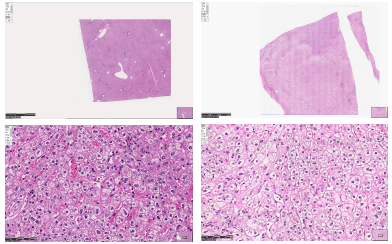
Chi-Square Goodness-of-Fit Numbers Obtained from Each Integer

Results

The second digit distribution of 500 cells' diameter did not adhere to benford's distribution (Graph 1) whereas the second digit distribution of the cells' area did adhere to Benford's distribution (Graph 2). The Pearson Chi-square goodness-of-fit test rejected the hypothesis for diameter but did accept the hypothesis for area.

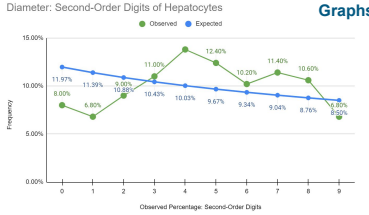
Discussion

The inconsistency between the area and diameter is due to the irregular shapes of the cells. The other limitation we experienced was attempting to automate the counting of the area using Image J. In the future, we hope to automate the process to test multiple cell lines in comparison to Benford's Law along with abnormal cells. We also hope to expand our research to include not only area but RGB and other methods used to contrast images.

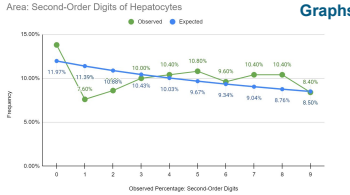

Normal hepatocytes

Metastatic Islet Cells

Graphs 1



Graphs 2



Conclusion

- Benford's application can be expanded to digital histology.
- This could be used as a screening tool with high sensitivity between natural and unnatural cells.
- More cells must be tested along with various cell lines to prove Benfordness throughout the body.

Reference

https://www.sciencedirect.com/science/article/pii/S1361841516301141
https://dl.acm.org/doi/abs/10.5555/2381219.2381243
https://pubs.geoscienceworld.org/esa/geology/article-abstract/40/4/327/130861/Applying-Benford-s-law-to-volcanoloxy?redirectedFrom=fulltext
https://books.google.com/books?hl=en&lr=&id=B2eYDwAAQBAJ&oi=fnd&pg=PR13&dq=%22benford%27s+law%22+&ots=RDpvV4aqvQ&sig=rLHkTfQnrisuB3DHRihF8nKJF0ifv#v=onepage&q=%22benford's%20law%22&f=false
https://office-watch.com/2021/benfords-law-and-excel/
https://www.journalofaccountancy.com/issues/2017/apr/excel-and-benfords-law-to-detect-fraud.html
https://bio-atlas.psu.edu/